wwPDB EM Validation Summary Report

Oct 14, 2024 – 04:56 PM JST

PDB ID	:	7F4G
EMDB ID	:	EMD-31450
Title	:	Structure of RPAP2-bound RNA polymerase II
Authors	:	Chen, X.; Qi, Y.; Wang, X.; Li, J.; Zhao, D.; Xu, Y.
Deposited on	:	2021-06-18
Resolution	:	2.78 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

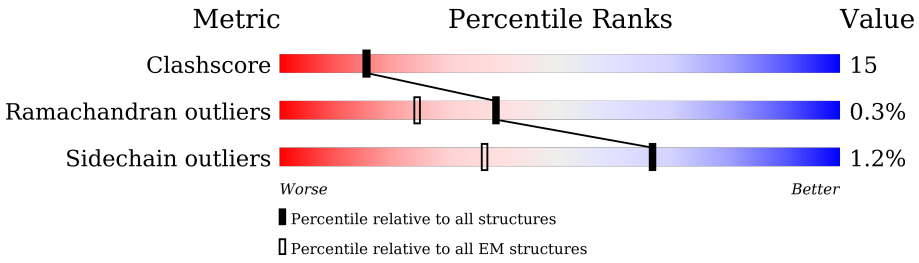
EMDB validation analysis	:	0.0.1.dev113
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 2.78 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



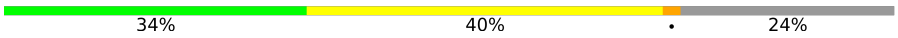
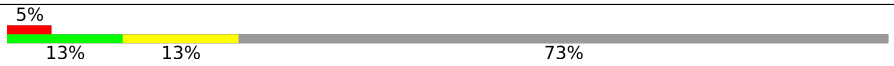
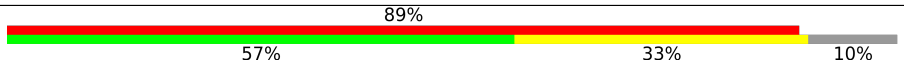
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1970	34% 18% 48%
2	B	1174	58% 25% 17%
3	C	275	67% 25% 8%
4	E	210	68% 31%
5	F	127	39% 20% 40%
6	H	150	77% 21%
7	I	125	69% 22% 9%
8	J	67	73% 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	K	117	
10	L	58	
11	R	612	
12	D	142	
13	G	172	

2 Entry composition (i)

There are 14 unique types of molecules in this entry. The entry contains 27957 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called RPB1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	1026	8179	5153	1440	1542	44	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	978	7820	4966	1356	1448	50	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	254	2038	1282	348	402	6	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II subunit E.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	E	208	1703	1080	299	316	8	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II subunit F.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	F	76	609	391	103	110	5	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	H	148	1186	750	194	237	5	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	I	114	Total	C	N	O	S	0	0
			927	571	166	179	11		

- Molecule 8 is a protein called RPB10.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	J	66	Total	C	N	O	S	0	0
			524	339	88	91	6		

- Molecule 9 is a protein called RNA_pol_L_2 domain-containing protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	K	115	Total	C	N	O	S	0	0
			920	593	152	173	2		

- Molecule 10 is a protein called RPB12.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	L	44	Total	C	N	O	S	0	0
			372	231	72	63	6		

- Molecule 11 is a protein called RPAP2.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	R	163	Total	C	N	O	S	0	0
			1335	848	227	253	7		

- Molecule 12 is a protein called DNA-directed RNA polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	D	128	Total	C	N	O	S	0	0
			1005	632	172	197	4		

- Molecule 13 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	G	171	Total	C	N	O	S	0	0
			1334	867	216	243	8		

- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

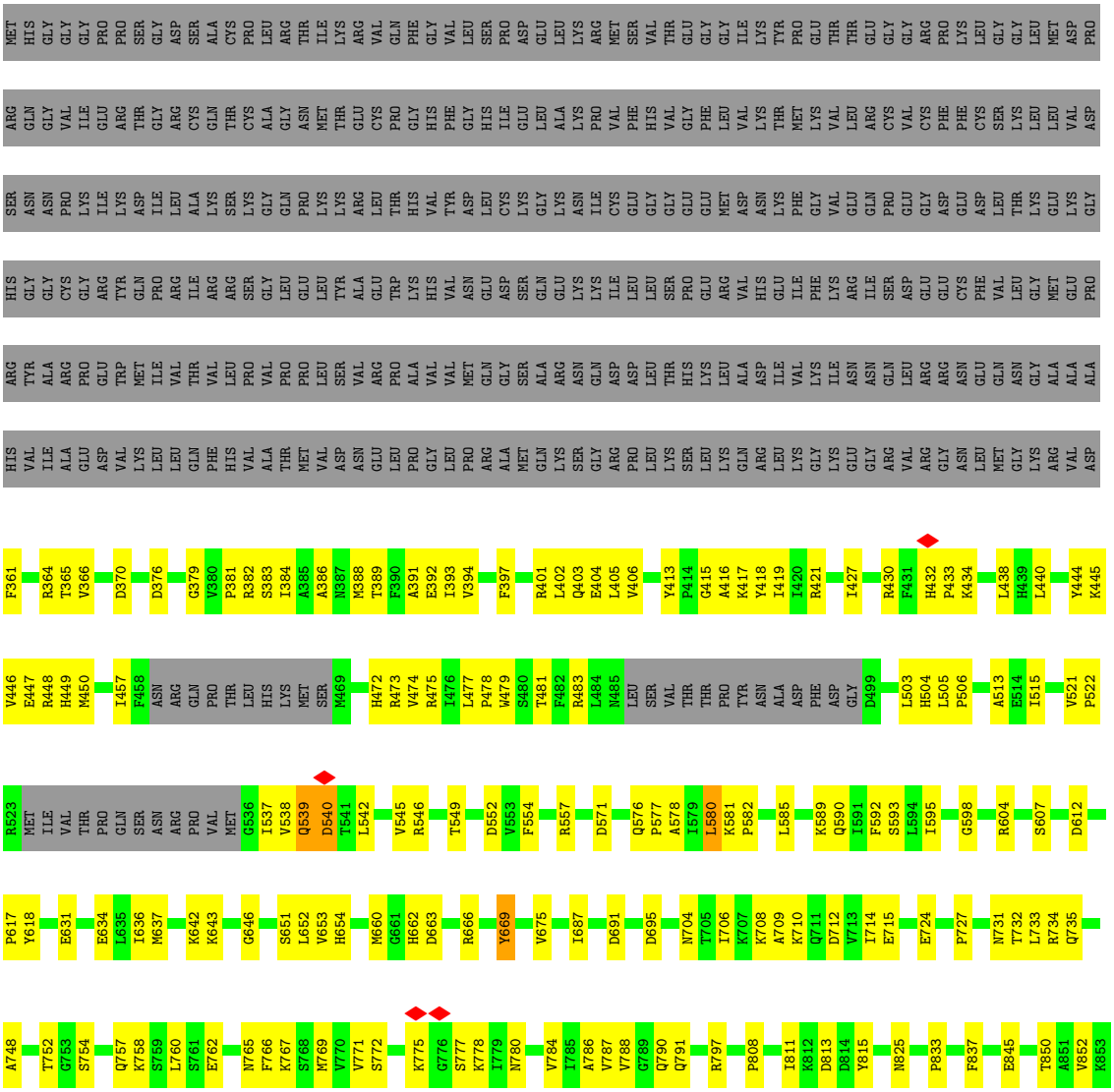
Mol	Chain	Residues	Atoms	AltConf
14	I	2	Total Zn 2 2	0
14	J	1	Total Zn 1 1	0
14	L	1	Total Zn 1 1	0
14	R	1	Total Zn 1 1	0

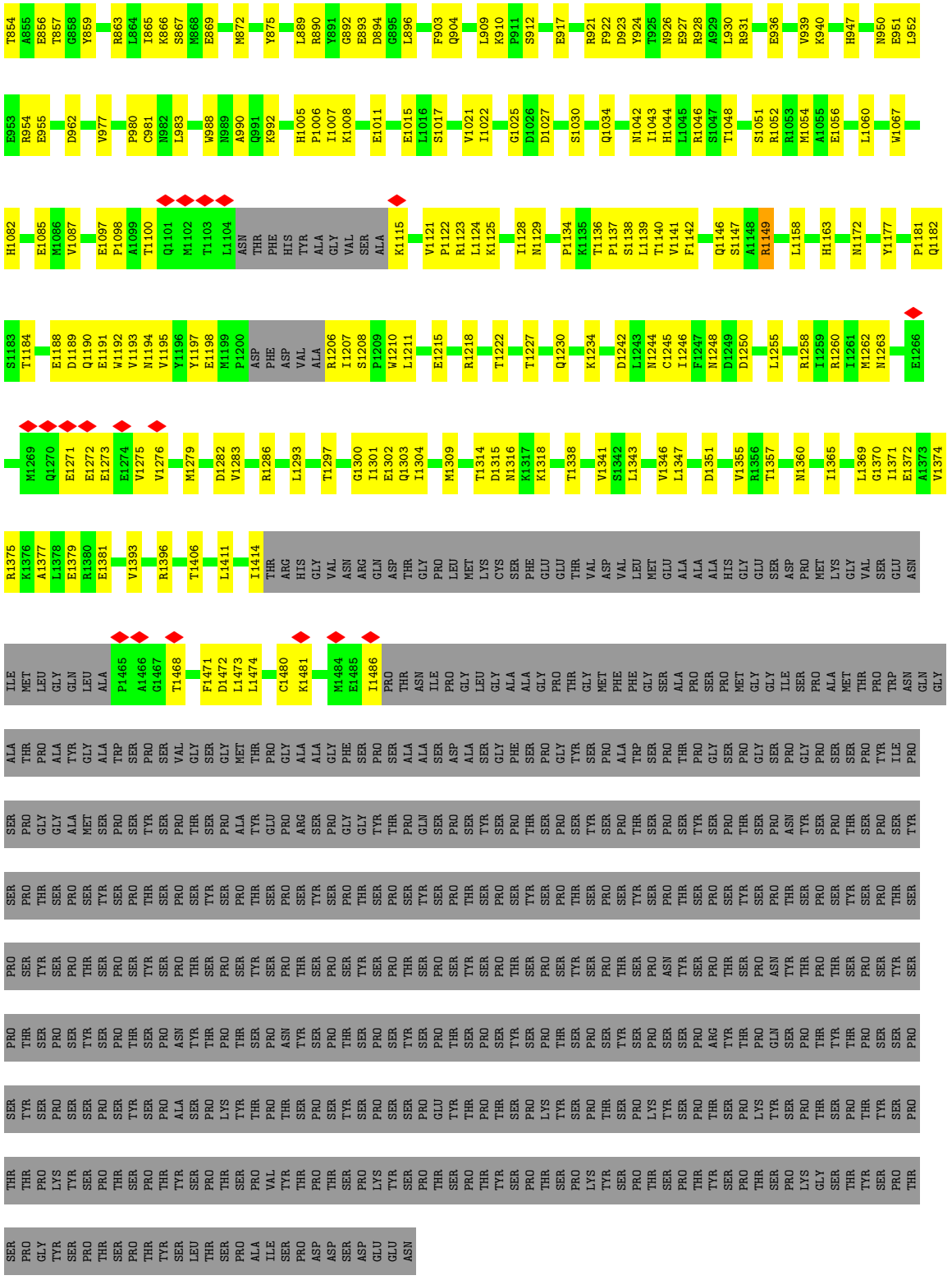
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: RPB1

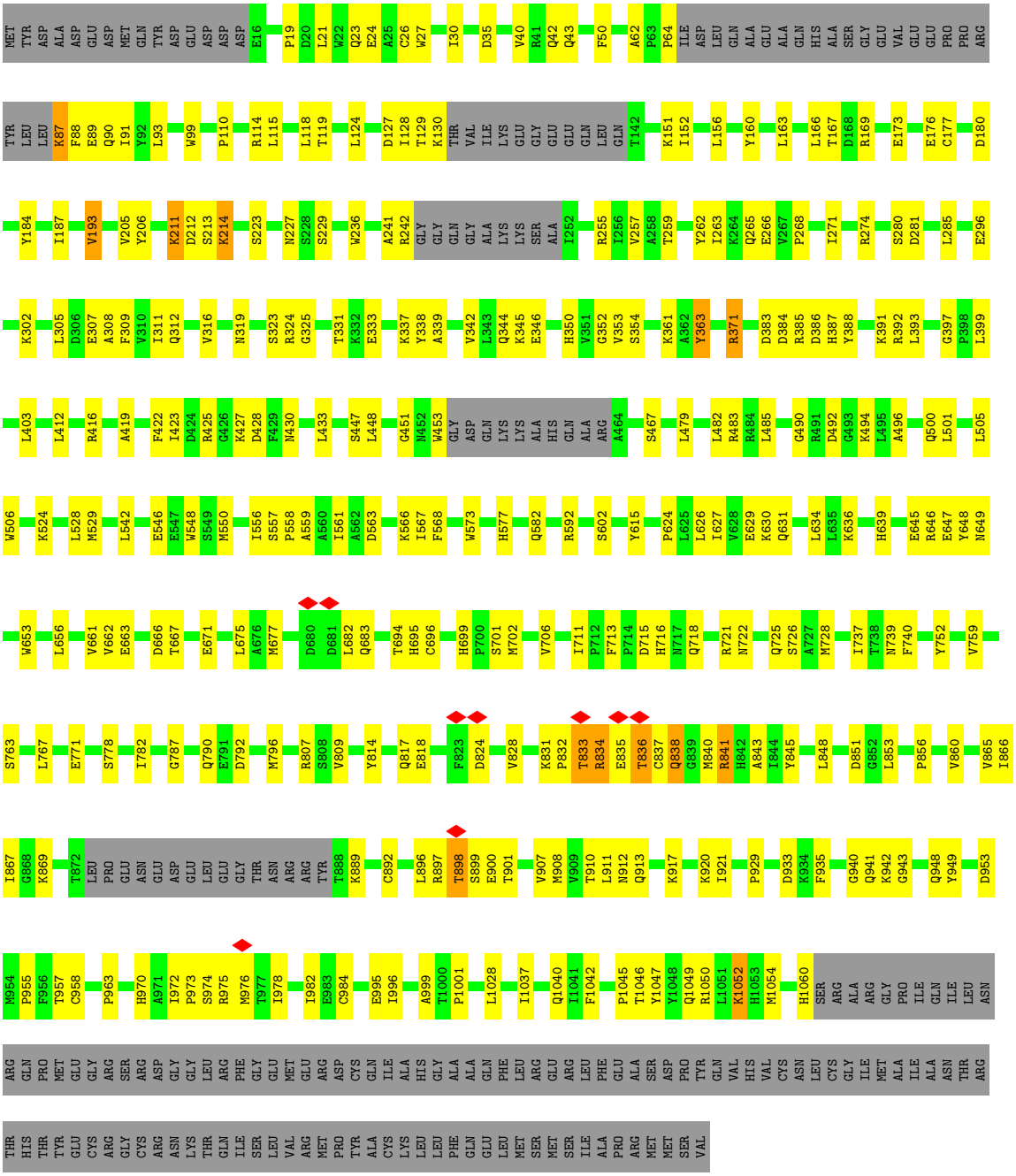
Chain A:





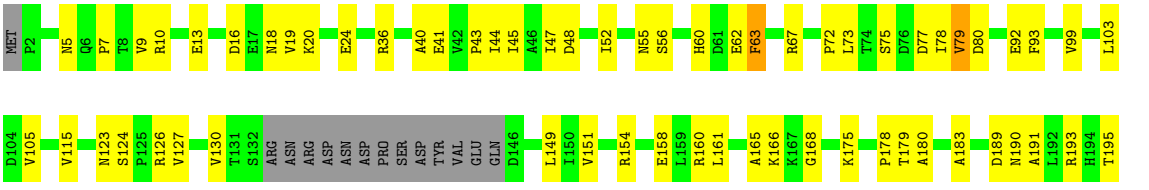
• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain B:



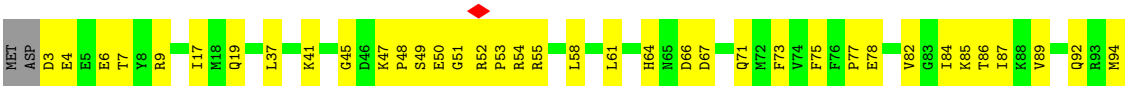
• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

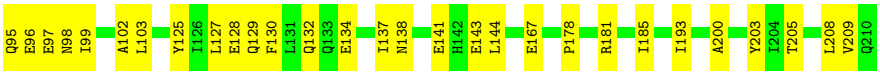
Chain C:



• Molecule 4: DNA-directed RNA polymerase II subunit E

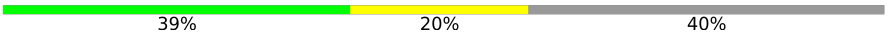
Chain E:





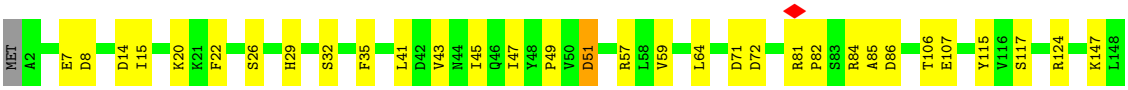
• Molecule 5: DNA-directed RNA polymerase II subunit F

Chain F:



• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:

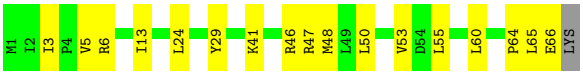


• Molecule 7: DNA-directed RNA polymerase II subunit RPB9

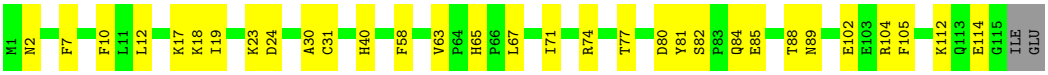
Chain I:

• Molecule 8: RPB10

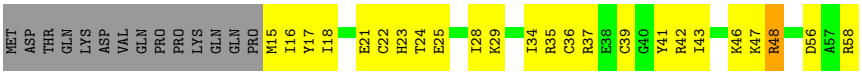
Chain J:



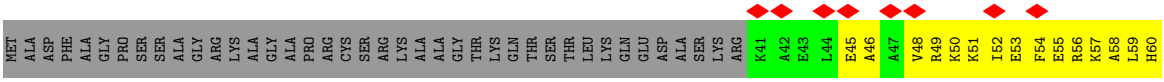
• Molecule 9: RNA_pol_L_2 domain-containing protein

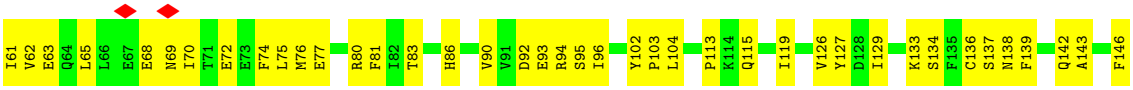


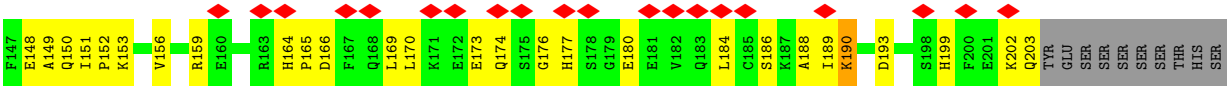
• Molecule 10: RPB12



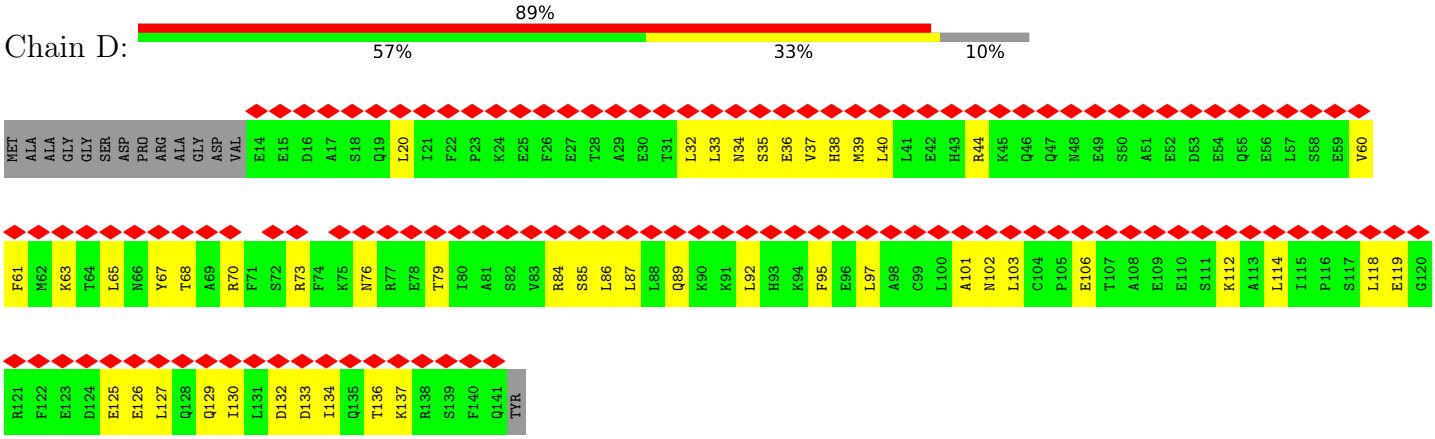
• Molecule 11: RPAP2



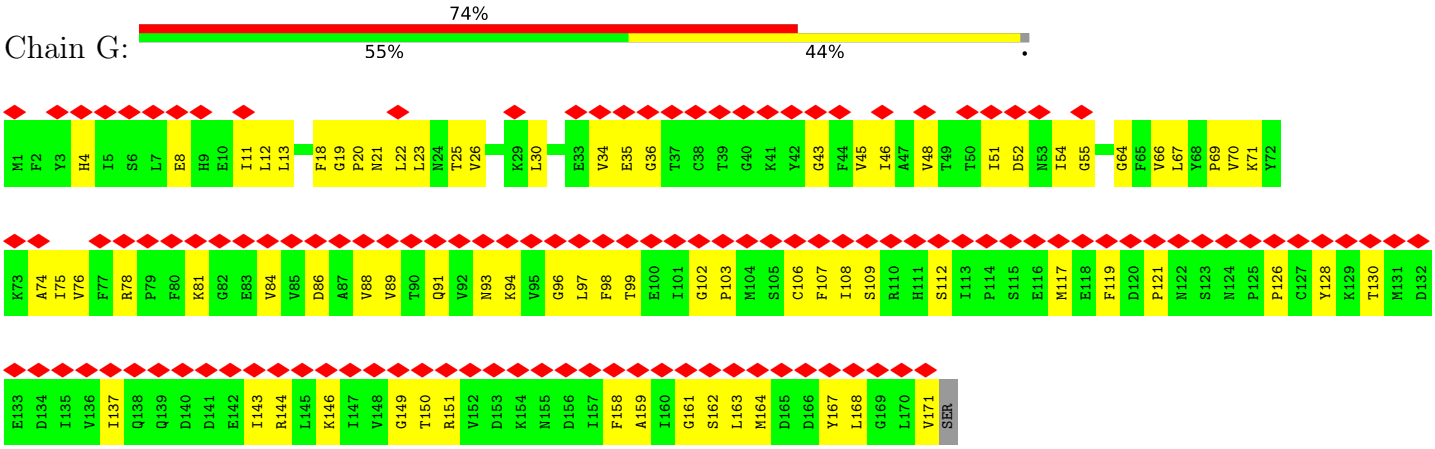




• Molecule 12: DNA-directed RNA polymerase II subunit RPB4



• Molecule 13: DNA-directed RNA polymerase II subunit RPB7



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	646517	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	5.185	Depositor
Minimum map value	-3.433	Depositor
Average map value	-0.003	Depositor
Map value standard deviation	0.111	Depositor
Recommended contour level	0.29	Depositor
Map size (Å)	334.08002, 334.08002, 334.08002	wwPDB
Map dimensions	320, 320, 320	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.044, 1.044, 1.044	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	A	0.68	0/8322	0.61	0/11237
2	B	0.92	2/7978 (0.0%)	0.66	2/10775 (0.0%)
3	C	1.01	0/2081	0.68	0/2828
4	E	0.61	0/1734	0.59	1/2342 (0.0%)
5	F	0.55	0/619	0.62	0/837
6	H	0.87	0/1207	0.60	0/1628
7	I	0.67	0/948	0.55	0/1284
8	J	1.08	0/533	0.76	0/719
9	K	0.82	0/939	0.63	0/1271
10	L	0.83	0/377	0.66	0/500
11	R	0.37	0/1362	0.52	0/1829
12	D	0.25	0/1019	0.50	0/1374
13	G	0.31	0/1365	0.53	0/1853
All	All	0.77	2/28484 (0.0%)	0.62	3/38477 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	193	VAL	CB-CG2	-5.78	1.40	1.52
2	B	1047	TYR	CD1-CE1	-5.49	1.31	1.39

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	556	ILE	C-N-CA	5.44	135.30	121.70
2	B	824	ASP	CB-CG-OD2	5.16	122.94	118.30
4	E	58	LEU	CA-CB-CG	5.16	127.17	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	8179	0	8216	278	0
2	B	7820	0	7823	208	0
3	C	2038	0	1991	52	0
4	E	1703	0	1725	52	0
5	F	609	0	640	24	0
6	H	1186	0	1147	23	0
7	I	927	0	860	24	0
8	J	524	0	540	14	0
9	K	920	0	942	23	0
10	L	372	0	382	22	0
11	R	1335	0	1329	70	0
12	D	1005	0	964	40	0
13	G	1334	0	1333	58	0
14	I	2	0	0	0	0
14	J	1	0	0	0	0
14	L	1	0	0	0	0
14	R	1	0	0	0	0
All	All	27957	0	27892	812	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 15.

The worst 5 of 812 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:721:ARG:HD2	2:B:975:ARG:HB3	1.48	0.94
2:B:387:HIS:NE2	2:B:671:GLU:OE2	2.05	0.88
2:B:953:ASP:OD1	3:C:36:ARG:NH2	2.07	0.88
1:A:542:LEU:HD21	1:A:642:LYS:HB3	1.56	0.87
11:R:94:ARG:HH12	11:R:103:PRO:HA	1.40	0.85

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	1012/1970 (51%)	916 (90%)	92 (9%)	4 (0%)	30 58
2	B	966/1174 (82%)	863 (89%)	98 (10%)	5 (0%)	25 52
3	C	250/275 (91%)	226 (90%)	24 (10%)	0	100 100
4	E	206/210 (98%)	195 (95%)	10 (5%)	1 (0%)	25 52
5	F	74/127 (58%)	70 (95%)	4 (5%)	0	100 100
6	H	146/150 (97%)	136 (93%)	10 (7%)	0	100 100
7	I	112/125 (90%)	99 (88%)	13 (12%)	0	100 100
8	J	64/67 (96%)	61 (95%)	3 (5%)	0	100 100
9	K	113/117 (97%)	104 (92%)	9 (8%)	0	100 100
10	L	42/58 (72%)	35 (83%)	6 (14%)	1 (2%)	5 15
11	R	161/612 (26%)	151 (94%)	10 (6%)	0	100 100
12	D	126/142 (89%)	119 (94%)	7 (6%)	0	100 100
13	G	169/172 (98%)	155 (92%)	13 (8%)	1 (1%)	22 48
All	All	3441/5199 (66%)	3130 (91%)	299 (9%)	12 (0%)	38 64

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	652	LEU
2	B	832	PRO
2	B	900	GLU
1	A	1302	GLU
2	B	19	PRO

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM

entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	910/1748 (52%)	897 (99%)	13 (1%)	62 85
2	B	857/1027 (83%)	838 (98%)	19 (2%)	47 76
3	C	231/252 (92%)	228 (99%)	3 (1%)	65 86
4	E	188/192 (98%)	188 (100%)	0	100 100
5	F	66/111 (60%)	66 (100%)	0	100 100
6	H	129/131 (98%)	128 (99%)	1 (1%)	79 92
7	I	103/112 (92%)	102 (99%)	1 (1%)	73 89
8	J	55/56 (98%)	55 (100%)	0	100 100
9	K	104/106 (98%)	104 (100%)	0	100 100
10	L	41/55 (74%)	41 (100%)	0	100 100
11	R	150/561 (27%)	149 (99%)	1 (1%)	81 93
12	D	106/126 (84%)	106 (100%)	0	100 100
13	G	147/153 (96%)	147 (100%)	0	100 100
All	All	3087/4630 (67%)	3049 (99%)	38 (1%)	66 87

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	841	ARG
6	H	51	ASP
2	B	898	THR
3	C	63	PHE
11	R	190	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 52 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	18	ASN
4	E	174	GLN
11	R	199	HIS
3	C	66	HIS
4	E	64	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

Of 5 ligands modelled in this entry, 5 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

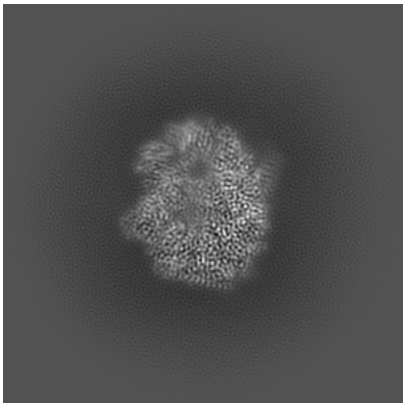
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-31450. These allow visual inspection of the internal detail of the map and identification of artifacts.

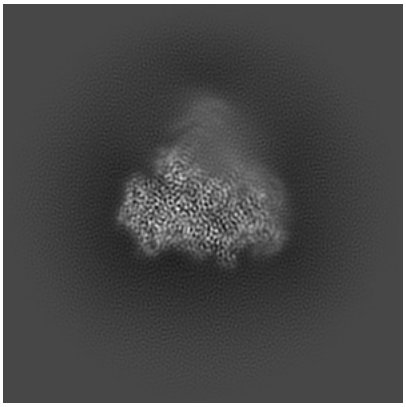
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

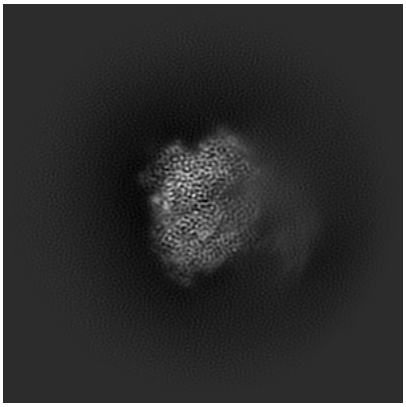
6.1.1 Primary map



X



Y

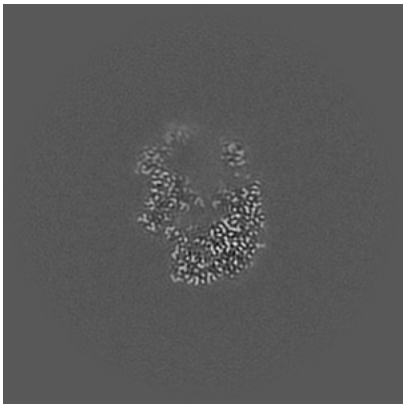


Z

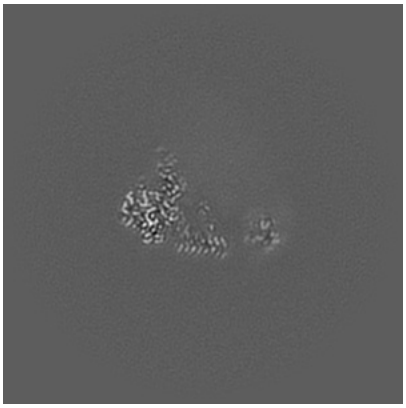
The images above show the map projected in three orthogonal directions.

6.2 Central slices

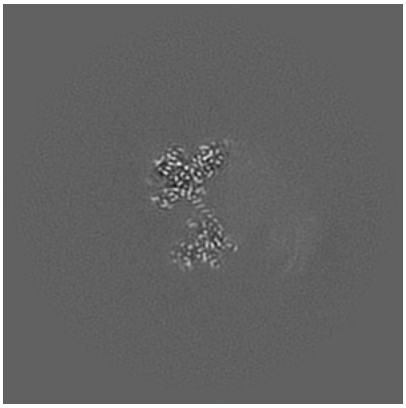
6.2.1 Primary map



X Index: 160



Y Index: 160

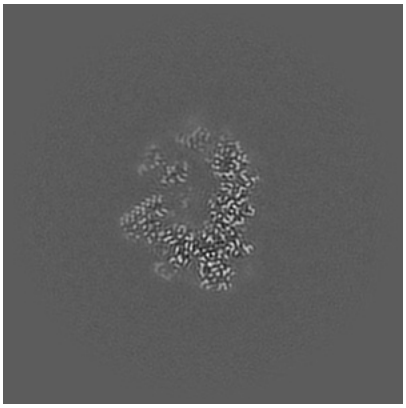


Z Index: 160

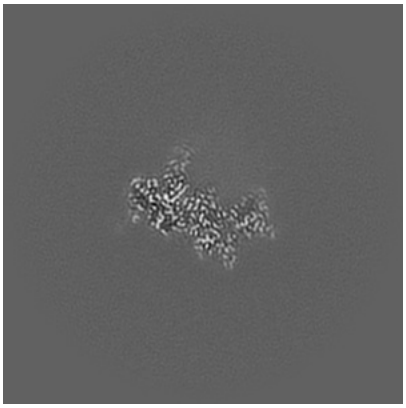
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

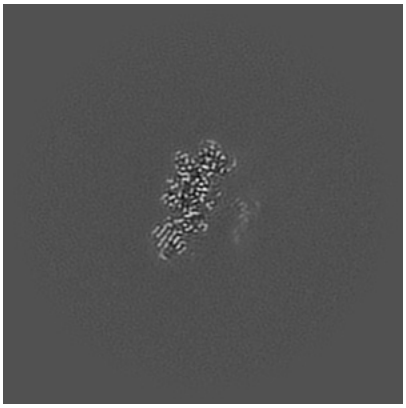
6.3.1 Primary map



X Index: 147



Y Index: 184

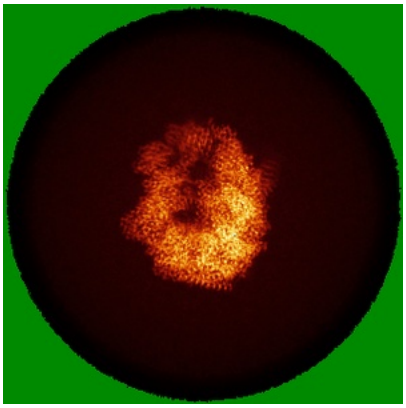


Z Index: 124

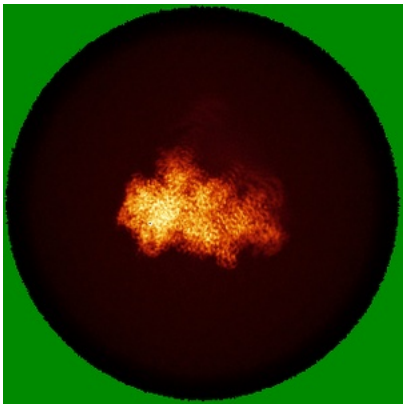
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) (i)

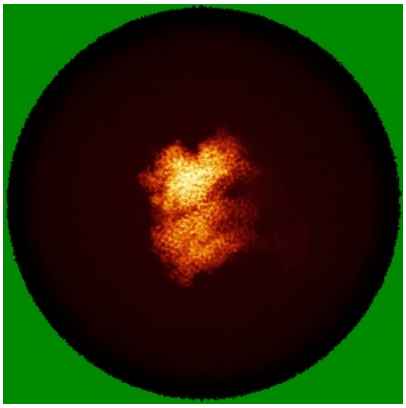
6.4.1 Primary map



X



Y

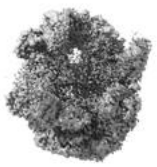


Z

The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

6.5.1 Primary map



X

Y

Z

The images above show the 3D surface view of the map at the recommended contour level 0.29. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

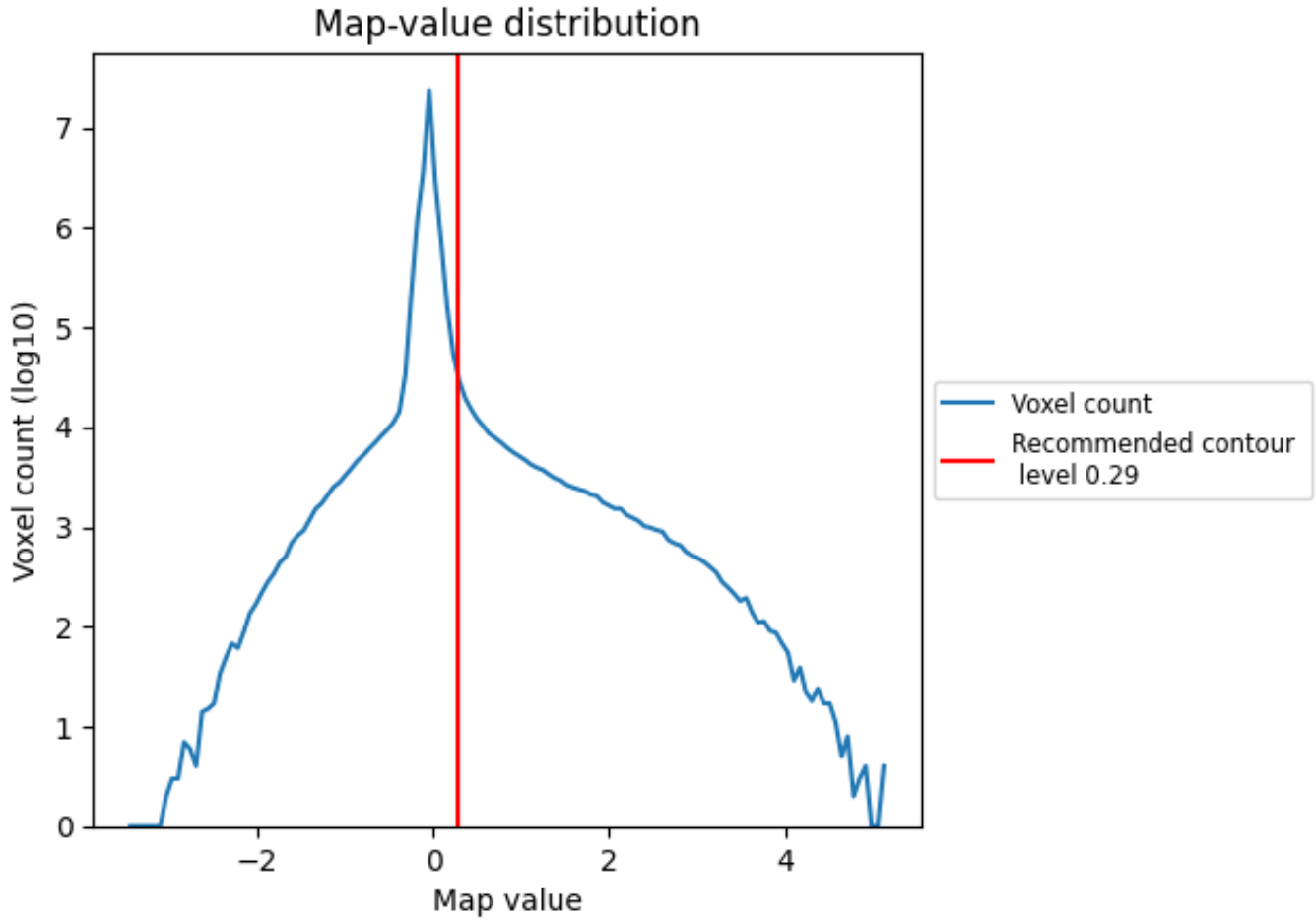
6.6 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

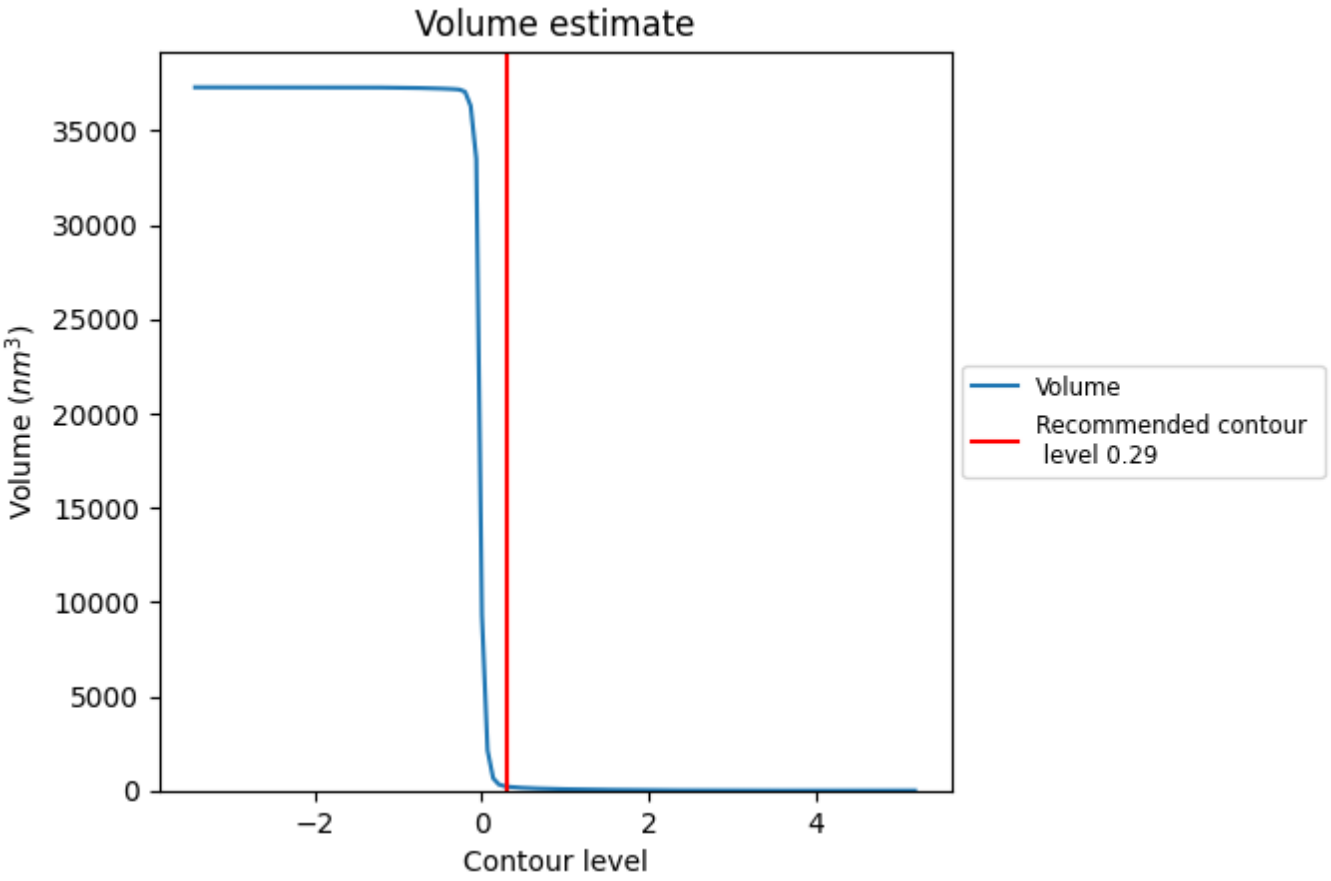
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

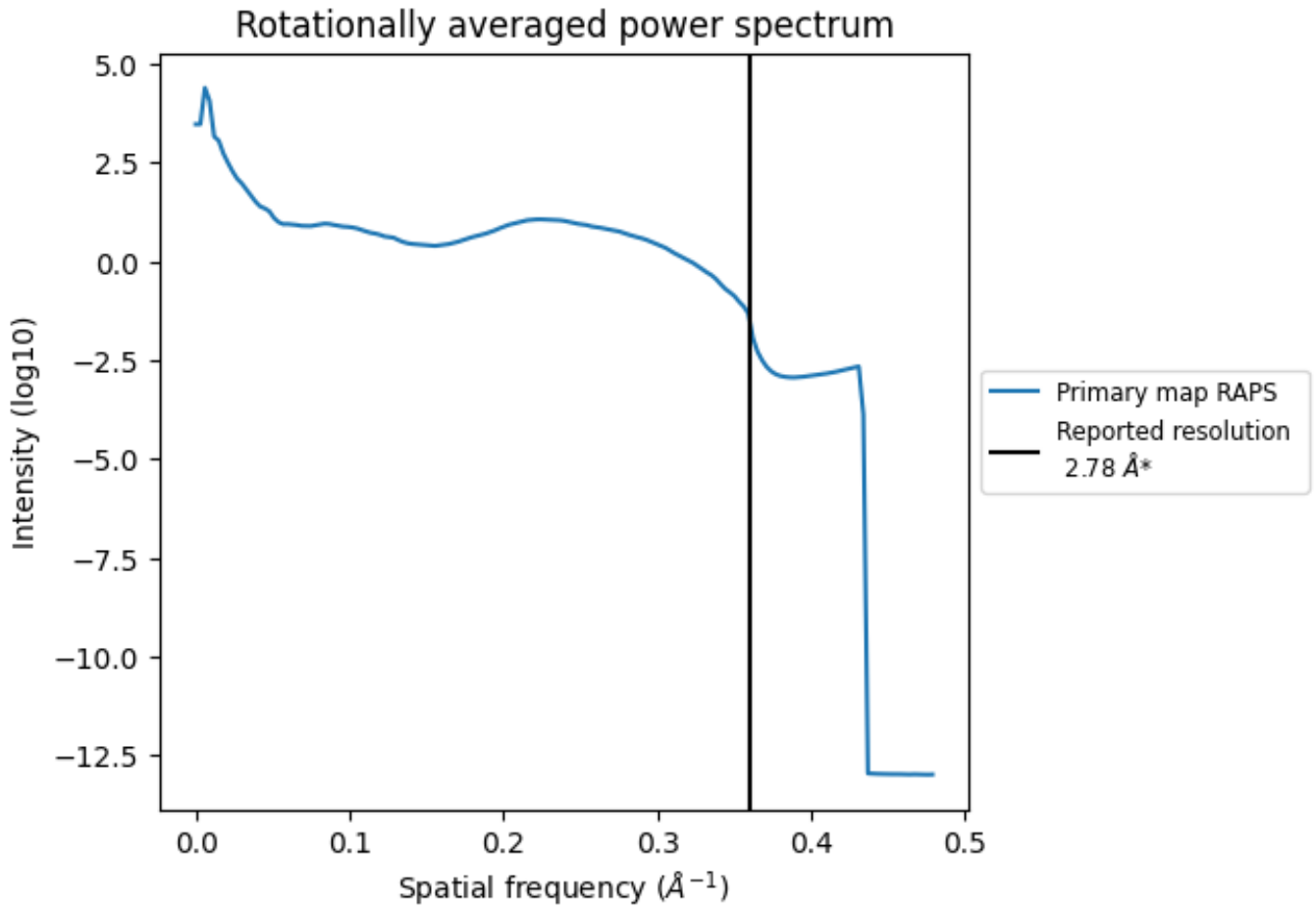
7.2 Volume estimate (i)



The volume at the recommended contour level is 222 nm^3; this corresponds to an approximate mass of 200 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum (i)



*Reported resolution corresponds to spatial frequency of 0.360 $Å^{-1}$

8 Fourier-Shell correlation (i)

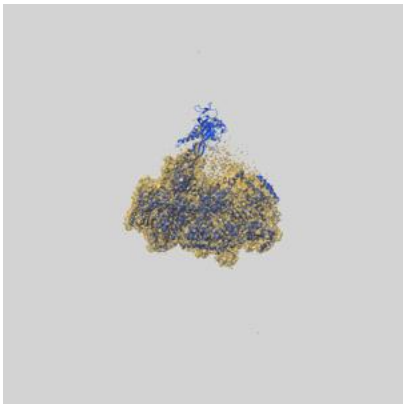
This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit

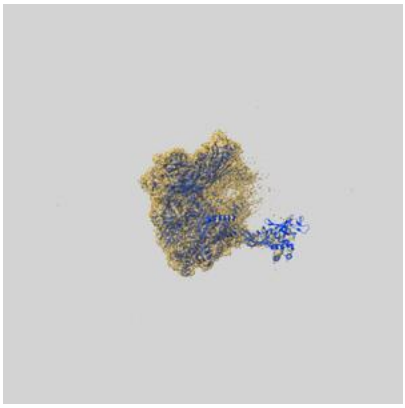
This section contains information regarding the fit between EMDB map EMD-31450 and PDB model 7F4G. Per-residue inclusion information can be found in section 3 on page 7.

9.1 Map-model overlay

X



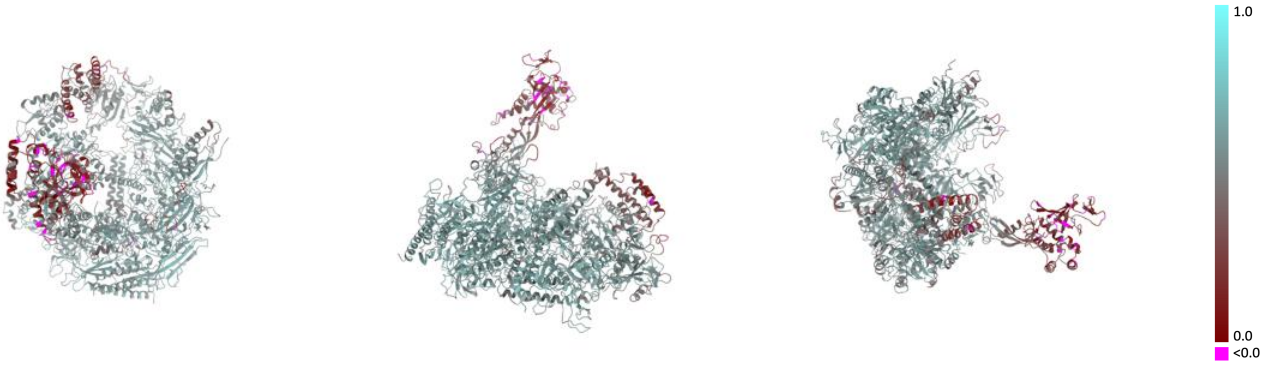
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.29 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

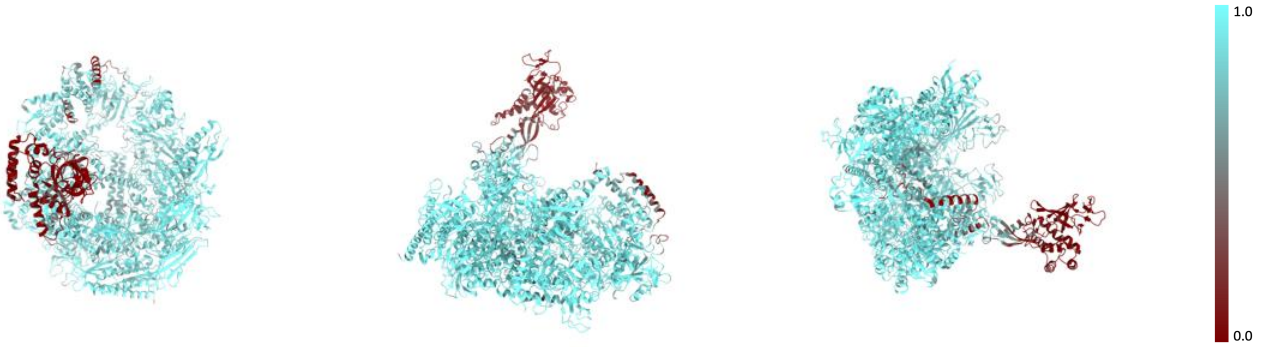
9.2 Q-score mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

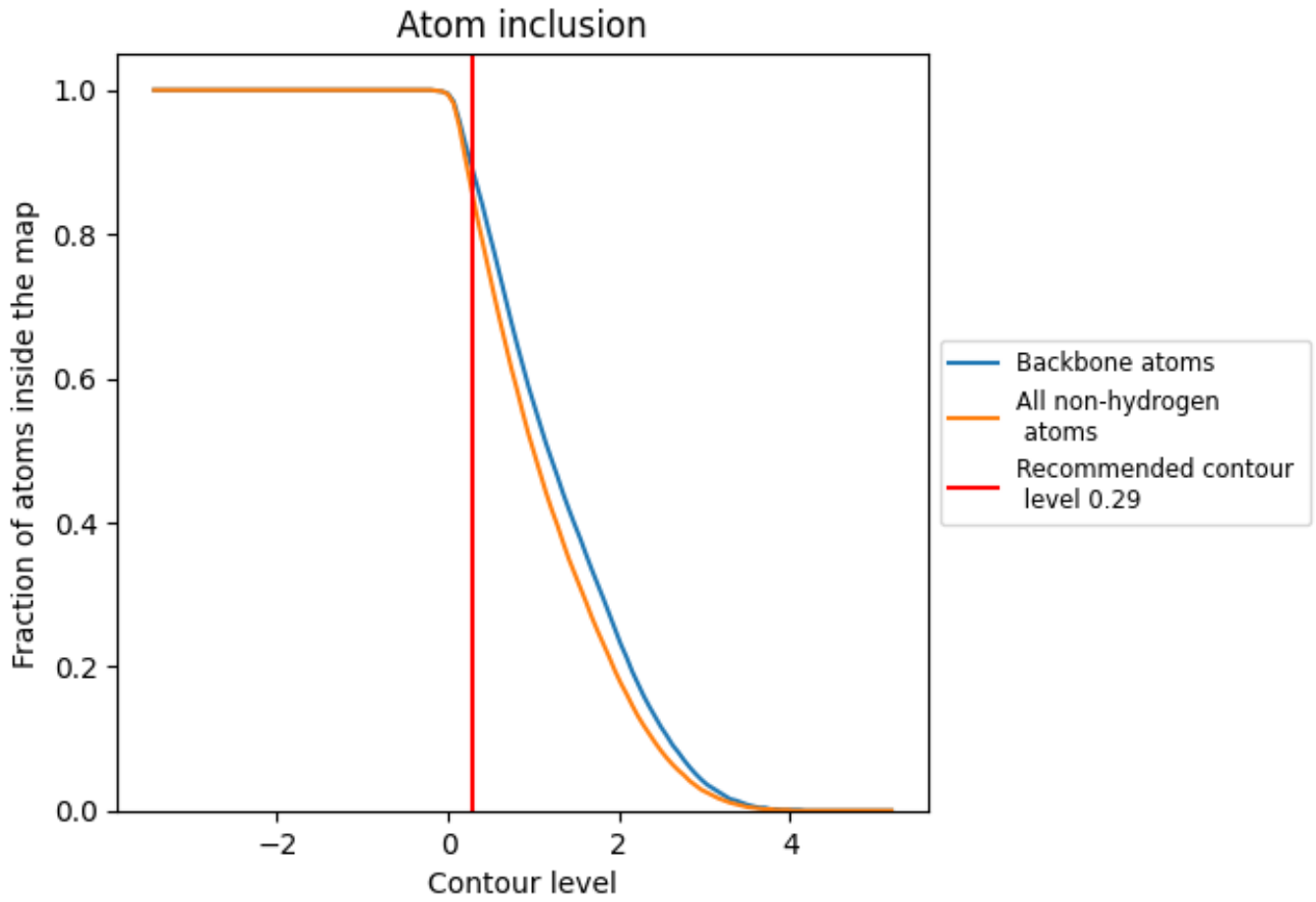
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.29).

9.4 Atom inclusion (i)



At the recommended contour level, 89% of all backbone atoms, 85% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary

The table lists the average atom inclusion at the recommended contour level (0.29) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8540	0.5300
A	0.9030	0.5520
B	0.9430	0.5780
C	0.9680	0.6000
D	0.0540	0.2140
E	0.9360	0.5480
F	0.8930	0.5370
G	0.2280	0.2600
H	0.9550	0.5770
I	0.9340	0.5550
J	0.9730	0.6190
K	0.9610	0.5930
L	0.9580	0.5510
R	0.6740	0.3370